*Patrizia Anesa, Ismael Arinas Pellón***

Disseminating and Hiding Information in Technical Communication Courses: The Case of Patents from the Perspective of Digital Humanities[1]

Abstract

Despite their ostensibly aseptic nature, technical texts involve a multifaceted net of institutional, social and pragmatic functions. Patents, in particular, are characterized by a multi-layered rhetorical exercise in which information is provided, and hidden, in light of patent disclosure laws.

Drawing on the Cooperative Patent Classification scheme, a corpus of patents related to environmental issues has been compiled. The objective is to investigate the main keywords emerging in the corpus and to analyse their semantic context in this patent type. More specifically, the analysis focuses on the patents' semantic preference and semantic prosody in order to pinpoint and examine the semantic complexities emerging in this genre and to identify the strategies employed (such as the use of linguistic vagueness) in order to provide the necessary information while not disclosing precious data.

Patents represent a complex, hybrid and cross-disciplinary genre and a finer understanding of their discursive features may contribute to spreading awareness of the importance that semantics plays within the rhetorical pattern of the text. Therefore, they may be fruitfully employed in Technical Communication courses in order to improved reading comprehension and analytical skills.

Keywords

environmental patents; keyness in patents; reading comprehension; intellectual property rights; knowledge disclosure; digital humanities

1. Introduction

Patents offer suitable tools for the development of reading comprehension skills in the field of technical communication. In particular, the complexity of a patent often lies in its dual nature as both a technical and a legal text, and the skills to be developed for its understanding are diverse and multifaceted. Burge (1999: 3-4) shows that even people who are frequently exposed to intellectual property (IP) issues may lack a profound knowledge of the differences between patents, trademarks, copyright and other intellectual property formats. Given the current importance of intellectual property rights (IPR) in assessing the performance of companies, institutions, and countries, it is pertinent to provide different stakeholders with the linguistic criteria needed in order to understand IPR-related genres and the textual patterns which characterize them. Thus, the patent application process lies at the crossroads of different areas of law and it is attracting growing attention in both business and academic settings.

1 This paper contributes to the national research program 'Knowledge dissemination across media in English: continuity and change in discourse strategies, ideologies, and epistemologies', funded by MIUR (PRIN 2015 no.2015TJ8ZAS). The paper has been planned and written jointly. However, Patrizia Anesa is responsible for sections 3 and 6, while Ismael Arinas Pellón is responsible for sections 2 and 4. Both authors are equally responsible for sections 1 and 5.

* Patrizia Anesa
University of Bergamo
E-mail: patrizia.anesa@unibg.it

Ismael Arinas Pellón,
Universidad Politecnica de Madrid
E-mail: ismael.arinas@upm.es

This study focuses on environmental patents.[2] The rationale behind this choice stems from the awareness that international initiatives on environmental issues have an impact on national laws and, in turn, these laws are setting trends in technological development. The commercialization of these technologies has quickly evolved into an extremely profitable market, consequently leading to further development beyond the original legal stimulus. In this respect, researching the language of environmental technology patents may serve to problematize the prototypical linguistic strategies deployed in this type of document.

This study contributes to the characterization of patents as a genre by providing strategies to access the explicit and tacit knowledge embedded in US environmental patent texts through keyword analysis (Adolphs 2006: 44; Bondi 2010: 4-5; Scott 2008: 176), semantic preference analysis, and semantic prosody analysis (see Hunston / Francis 2000: 104-106; Partington 1998: 66-69; Stubbs 2002: 65-66; Stubbs 2009: 125). The applicability of the information obtained to didactic processes and curriculum development is also discussed. In particular, US patents may be used as sources of specialized information to develop the reading comprehension skills of, for example, Law and Engineering undergraduates. More specifically, we focus on gaining a finer understanding of how language is subject to a genre-specific interpretation when describing intellectual property and introducing the need for a patented invention.

Thus, our goal is to observe how the analysis of patents can be employed to improve reading comprehension and our aims focus on two main areas:

1) which factors affect the drafting of a document, for example legislation,[3] document-drafting purposes (e.g. getting IP, commercial profit, increasing the value of a company, etc.), and any potential reaction of the readership;[4]

2) how information is explicitly presented or intentionally obscured.

This paper focuses, firstly, on the analysis of common words (through keywords) which structure patent claims, in order to observe their specialized semantics in this genre. Subsequently, we observe the role that vague nouns (again keywords) typically play in describing inventions. In this respect, we also describe how the use of *-able/-ible* adjectives is carefully chosen by lawyers to cover as much IP scope as possible. Finally, we reflect on how the linguistic analysis of these documents can be implemented in Technical Communication courses.

2. Patents as a genre

Patents in the US are granted for a period of 20 years from their application date. The following items are defined as protectable inventions: designs, plants, compositions of matter, methods, and devices.[5]

Rhetorically, US patent applications must be convincing in terms of novelty (35 USC[6] §102), non-obviousness (35 USC §103), and utility (35 USC §101). Novelty results from clearly differentiating the patentable invention from existing prior art. Non-obviousness refers to the

2 Unless otherwise specified, in this study the term 'patent' is used to refer to the patent document (rather than the right that it generates).

3 Manual of Patent Examining Procedures (MPEP), and US Code: Title 35 – Patents, available at: http://uscode.house.gov/browse/prelim@title35&edition=prelim (accessed July 7, 2018).

4 Rejection or granting of patent by examiners; being or not being sued for patent infringement by competitors; court rulings; imposing technical standards.

5 Design patents in many countries are protected under a different category: industrial designs. However, industrial design protection is granted by the patent offices of those countries making this distinction. Plant patents are one peculiarity of the US intellectual protection system. Under the category of compositions of matter, some living tissue can be patented only in the US.

6 See US Code: Title 35 – Patents.

requirement for an invention not to be an apparent variation of a prior one. The utility criteria is fulfilled if some functional or technical gap is identified in previously patented, related inventions.

The information contained in patents is shaped for a restricted community of practice (Wenger et al. 2002: 4). Those scientists, engineers and lawyers outside the patent community will find it problematic to access this specialized information because of their unfamiliarity with the conventions followed within this specific community (Myers 1995). To understand patents, readers must be aware of their explicit and implicit technical knowledge, as well as the legal practices that guide the word choices displayed (Ewing et al. 2014). In particular, the understanding of the complex information dealt with in a patent needs complex hermeneutic processes. Patents indeed represent a genre which involves the technological, scientific, financial, and legal spheres and, thus is inherently interdisciplinary (Brugnoli 2007). Consequently, processing the knowledge contained in patents demands mastering a wide range of competences. In this respect, this study offers considerations on the codification processes implied in the specific mode of communication of knowledge in US patents. The focus is on several linguistic conventions used in patent documents both to describe the property claimed as well as the usefulness, novelty and non-obviousness of the invention.

Essentially, in US patents the written disclosure of the invention consists of three parts: Field and Background of the Invention, Description, and Claims. The first two have technical implications, while the third lays out the scope of the property protected by the patent. The claims constitute the specific[7] section in US patents containing vocabulary with particular legal construal (see section 5.2 below). There are two types of claims: independent and dependent. The former defines one or more aspects of the protected invention, while the latter introduces limitations to one or more independent claims (Slusky 2007: 105-106). In section 5.2, we present the vocabulary used either in dependent or independent claims to limit the scope of the inventions.

3. The case of environment-related patents

The Rio de Janeiro Earth Summit[8] (held in June 1992) and the adoption of the Kyoto Protocol in December 1997 triggered a series of changes in environmental legislation which have influenced the development of a host of technological fields. More recently, the significances of subsequent international initiatives such as the Copenhagen Climate Change Conference (December 2009) or the Paris Agreement on Climate Change[9] (November-December 2015) have further influenced environmental laws and technological developments. Patents are, as a result, a reliable reflection of these environmental trends. For example, following these documents, the European Union has established policies and directives promoting greater energy efficiency, the development of competitive clean-energy technologies, the reduction of polluting emissions, the control of the use of harmful chemicals, and the reduction of levels of noise pollution.[10]

The growth of patent innovations has been dramatic in recent years. This trend regards all inventions, but in particular patents connected to environmental technology. Whether patents can successfully promote or channel innovation remains subject to lively debate (Burk / Reyman 2014). However, unlike other forms of intellectual properties, their rhetorical structure of patents is embedded in specific legal needs and compositional constraints, which determine a certain level of linguistic conservatism. Indeed, patents have to follow precise canons in that they are

7 Other words defined in the US patent glossary (https://www.uspto.gov/learning-and-resources/glossary) may appear in any section of the patent specification.

8 United Nations Conference on Environment and Development (UNCED), Rio de Janeiro, 3-14 June 1992.

9 See *Paris Agreement* available at:
http://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf (accessed January 20, 2017).

10 See Health and Wellbeing, European Commission, available at: http://ec.europa.eu/environment/basics/health-wellbeing/index_en.htm (accessed January 20, 2017).

subject to a review conducted by a governmental agency and their coming into existence depends on a formal act of government (Burk / Reyman 2014: 168).

The social role played by patents is evident in the knowledge society era. Given the growing importance of "greening discourse" in different contexts, environmental patents need scholarly investigation in order to understand their relevance not only as a metaphor or an evaluative tool of technological innovation but also as a fundamental communication means used by companies, institutions, and individuals.

4. Methodology

4.1 Corpus

We compiled a corpus of 111 patents dealing with environmental categories (from now on referred to as 1MEPAC). This corpus holds approximately 1 million tokens. The document selection was conducted on the basis of the patent classification scheme developed jointly by the United States Patent and Trademark Office and the European Patent Office, namely the Cooperative Patent Classification (CPC).[11]

1MEPAC includes patents classified under the following nine general categories:[12]

1. Y02B 10/00 Integration of renewable energy sources in buildings.
2. Y02B 20/00 Energy efficient lighting technologies.
3. Y02B 30/00 Energy efficient heating, ventilation or air conditioning (HVAC).
4. Y02B 40/00 Technologies aiming at improving the efficiency of home appliances.
5. Y02B 50/00 Energy efficient technologies in elevators, escalators and moving walkways.
6. Y02B 60/00 Information and communication technologies (ICT) aiming at the reduction of own energy use.
7. Y02B 70/00 Technologies for an efficient end-user side electric power management and consumption.
8. Y02B 80/00 Architectural or constructional elements improving the thermal performance of buildings.
9. Y02B 90/00 Enabling technologies or technologies with a potential or indirect contribution to Green-House Gas emissions mitigation.

The compilation was based on a random selection of 12 recent patents in each category. The corpus details are illustrated in Table 1.

11 More specifically, the classification scheme "Y2" dealing with climate-change mitigation technologies. This scheme can be consulted at https://www.cooperativepatentclassification.org/cpc/scheme/Y/scheme-Y.pdf (accessed May 20, 2018).

12 Available at: http://www.cooperativepatentclassification.org/cpc/scheme/Y/scheme-Y02B.pdf (accessed January 11, 2017).

Number of US patents in corpus	111
Years covered by corpus	2011-2016
Tokens excluding non-alphabetic characters	1,078,639
Types excluding non-alphabetic characters	15,897
Type/Token Ratio (TTR)	14.6%
Token/Type Ratio	67.8
Average tokens per document	9,717.46
Hapax Legomena (% of total Types)	4,803 (30%)
Number of keywords extracted	300

Table 1. Environmental Patent Corpus Description

4.2 Approach

Rather than on word frequencies, our focus was specifically on keywords.[13] According to Scott (2008: 176), keywords are words "whose frequency is unusually high in comparison with some norm". Moreover, Goźdź-Roszkowski (2011: 35) states that keywords can also "reveal not only a great deal about the subject matter, the 'aboutness' of a particular genre, but they can also specify the salient features which are functionally related to the genre". Therefore, our initial strategy consisted of extracting two types of keywords from the 1MEPAC corpus: (a) keywords which reflect prototypical linguistic characteristics of patents as a genre; and (b) keywords that reflect the possible differences between patents in general and patents dealing with inventions related to the environment. For the first purpose, we compared 1MEPAC to the 100 million tokens of the British National Corpus (BNC). For the second purpose, we used a 7-million-token US patent corpus previously compiled for researching patents as a genre (from now on 7MUSPAC).[14] Thus, both the BNC and the 7MUSPAC provide a standard against which to compare the 1MEPAC for our analysis (cf. Rayson and Garside 2000). The results of these comparisons are displayed in Table 2 (Section 5.1), and for practical reasons only the first 20 keywords from each comparison are considered. The analysis was conducted using the software package Antconc 3.4.4w,[15] which offers two statistical tests to detect keywords in a corpus: the Chi-squared test and the Log-likelihood test. Rayson and Garside (2000) and Rayson, Berridge, and Francis (2004: 4) recommend the Log-likelihood test on the grounds that word-types do not appear in a normal distribution across different texts in a corpus.

> For our analysis, we checked the collocation, colligation, semantic preference, and semantic prosody of a selection of the highest-ranking words in the keyword list. Additionally, we analysed adjectives ending in *-able* and *-ible* as some patent attorneys and translators directly (Cole 2007: 160; Slusky 2007: 95; Erwing 2014: 94) or indirectly (Lawson 1997: 176; Roberts 2007: 85-86) indicate that an inadequate choice of adjectives may give an advantage to competitors (see Section 5.3). To identify these adjectives in the 1MEPAC, we Part-Of-Speech tagged it using TXM,[16] which also allows the searching of tagged corpora.

4.3 Working definitions

As collocation, colligation, semantic preference, and semantic prosody represent the privileged object of this analysis, the interpretation of these concepts as adopted in this study will now be

13 We have considered the frequency per 10,000 words as this is the average length of the patents in our corpus.
14 This corpus contains 1,001 patents spanning the years 1995 to 2018.
15 AntConc is available at: http://www.laurenceanthony.net/software.html (accessed January 12, 2017).
16 TXM is available at: http://textometrie.ens-lyon.fr/spip.php?rubrique96&lang=en (accessed January 19, 2018).

presented and exemplified through excerpts from our corpus. Stubbs (2002: 64-66; 2009: 125) defines these four patterning parameters as follows: *Collocation*: the usual co-occurrence of words. For example, in this corpus the forms of the verb *overcome* tend to be followed by words such as *problems*, *disadvantages*, *deficiencies*, *limitation*, or *issues*, as Figure 1 illustrates:

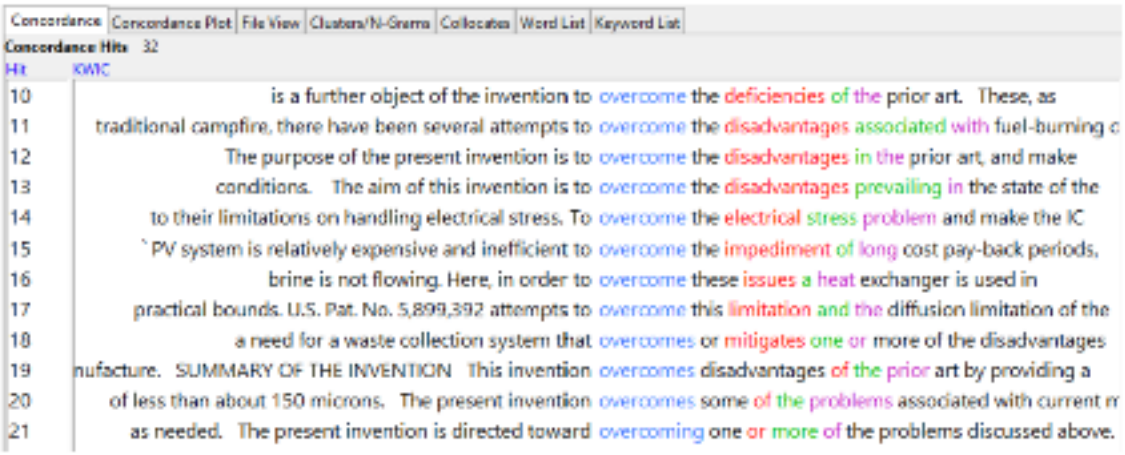

Figure 1. Example of Collocation with *overcome.*

- *Colligation*: the usual co-occurrence of grammatical choices. For example, the word *may* tends to be followed by the adverb *further* and a verb or a passive form or both, as shown in Figure 2:

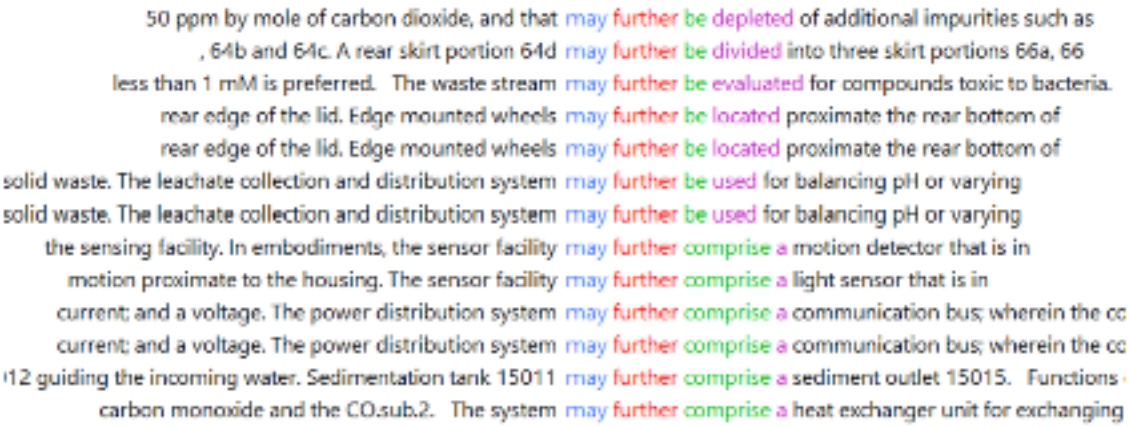

Figure 2. Examples of Colligations with *may.*

- *Semantic preference:* the semantic grouping of the words which co-occur on either side of the node (keyword). For example, battery / batteries occur 3,384 times in our corpus and over one third of those occurrences appear in the context of technical characteristics and operation with embedded (467 times), powered (425 times), rechargeable (306 times), pack (239 times), or charge (190 times).

- *Semantic prosody*: the connotations in the context of a lexical item, not easily discovered solely by intuition. These connotations are described in terms of positive or negative polarity, tentativeness, indirectness, and face-saving associations. For example, in our corpus the word *waste* always appears in a context where it is submitted to a process, thus presenting it from a neutral standpoint. Instead, the word *cost* appears in contexts of saving, reducing or efficiency, therefore suggesting an advantage of the described invention.

5. Analysis and discussion

5.1. Keywords

Table 2 presents the 20 highest-ranking keywords (excluding prepositions, articles, plural forms and conjunctions) in our corpus. The absolute frequency in the corpus is provided in brackets.

Using BNC as Reference Corpus (Group A)		**Using 7MUSPAC as Reference Corpus (Group B)**	
Absolute frequency	**/10,000 words**	**Absolute frequency**	**/10,000 words**
Power (6,937)	64	Energy (4,906)	45
Energy (4,906)	45	Power (6,937)	64
Embodiment (2,898)	27	Lighting (1,763)	16
Invention (3,128)	29	Water (4,449)	41
Device (2,887)	27	May (10,597)	98
Battery (2,609)	24	Waste (1,777)	16
Water (4,449)	41	Bulb (1,482)	13
Heat (2,987)	28	Light (3,880)	36
Light (3,880)	36	Wireless (1,920)	18
Wherein (1,916)	18	Battery (2,609)	24
System (4,108)	38	Solar (1,326)	12
Wireless (1,920)	18	Heat (2,987)	28
Gas (2,650)	25	Carbon (1,675)	16
Temperature (2,379)	22	AC (1,133)	11
Fluid (1,970)	18	Source (2,612)	24
Source (2,612)	24	LED (919)	9
Lighting (1,763)	16	Grid (767)	7
Control (2,933)	27	CO[17] (1,777)	11
Example (3,041)	28	Building (917)	9
Thermal (1,442)	13	Batteries (775)	7

Table 2. Selected Corpus Keywords (ordered by keyness)[18]

Table 2[19] shows the keywords in our corpus in relation to, firstly, general vocabulary (Group A) and, secondly, the vocabulary of other patents (Group B). In the first case, approximately one third of these keywords are either typical section headings, phraseology introducing details regarding the features of the invention, or formulaic language which stresses the legal consequences of infringing the IP protected by the claims (Arinas Pellón 2010: 321; Arinas Pellón 2012). In the second case, with the exception of the modal verb *may*, all the keywords are related to technologies whose environmental-friendliness is being improved.

We analyse, first, the special meaning that selected keywords from Group A and Group B have in patents. Several of these keywords are associated with other words whose construal in patents plays an essential role in claiming property. Then, we observe the collocates of *-able/-ible* adjectives, which were not detected with the keyword approach, but whose reading have

17 Including CO and CO_2.

18 Chi-Squared.

19 The three acronyms stand for Alternating Current (AC), Light Emitting Diode (LED), and Carbon Dioxide or Carbon Monoxide (CO).

important implications for what the patented property is (Arinas Pellón 2014: 265). The section concludes with a semantic preference and prosody analysis of selected keywords from Group B in Table 2. This analysis is limited to the patent section generally known as "Background of the Invention" where arguments for the usefulness, novelty and non-obviousness of the patent are developed.

5.2. Keywords with special meaning in US patents

Looking at the list of keywords which result from the comparison of our corpus with the BNC[20] (Table 2), one may have the initial impression that it does not display any words with a specifically "legal" meaning. The collocates of five of these keywords (*invention*, *embodiment*, *wherein*, *comprising*, and *may*) either have a special meaning in patents or are typically used in patents to achieve the goals of patentees.

In particular, both *embodiment* (2,898) / *embodiments* (2,881) and *invention* (3,131) / *inventions* (10) are considered significant enough by the USPTO to include them in the glossary[21] compiled for patent applicants. They define *invention* as "any art or process (way of doing or making things), machine, manufacture, design, or composition of matter, or any new and useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States". *Embodiment* is defined in the same glossary as "a manner in which an invention can be made, used, practiced or expressed".

Table 3 displays a selection of the most frequent collocations of invention. Those in the left-hand column introduce specific features of the invention, although patent drafters appear to prefer using the word embodiment for this function (see Table 4 below). Those in the right-hand column are used either as headings of document sections (Scope of the Invention, Field of the Invention, Background of the Invention, etc.) or to introduce lists of features as example (1) illustrates.

Distinctiveness of the Invention	**Patent Sections and Feature Lists**
Object(s) of the [present] invention (62) Advantage(s) of the [present] invention (36) Feature(s) and advantage(s) of the [present] invention (14) Applicability of the invention (6) [Further / Other] objects and advantages of this invention (4)	Scope of the invention (92) Aspect of the invention (59) Summary of the invention (59) Field of the invention (57) Background of the invention (52) Detailed description of the invention (37) System of the invention (30) Method of the invention (25) Container of the invention (15) Principles of the invention (15) Spirit of the invention (15) Article of the invention (14) Object of the invention (14) Practice of the invention (14) Aspects of the invention (13) Cap of the invention (12) Labels of the invention (12) Advantages of the invention (11)
Total: 124 occurrences	Total: 1,055 occurrences

Table 3. Selected Collocations for *Invention*

20 As illustrated in Section 4.2, the choice to conduct the first part of the analysis using the BNC as a reference corpus derives from the need to employ a well-established corpus of General English and, secondly, from practical reasons related to the availability of the corpus. Potential discrepancies deriving from differences between geographical varieties will be signaled, if present.

21 The glossary is available at: https://www.uspto.gov/learning-and-resources/glossary (accessed March 28, 2019).

(1) *Water treated by a method or* ***system of the invention*** *can be freshwater, brackish water, or seawater*. (U.S. Patent 9,051,153)

Table 4 displays a selection of collocations with the keyword *embodiment*. In this case, the collocations on the left are used to present possible implementations of the patented invention, as in example (2), and the collocations on the right to describe the characteristics of these implementations (example 3).

Alternative Forms of the Invention	**Collection of Features of the Invention**
In any of the above embodiments (22)	[...] embodiment of the invention (200)
In [yet] an alternate embodiment (102)	[...] embodiment of the present invention (265)
In alternate embodiments (195)	[...] embodiments of the invention (191)
Alternate embodiments (15)	[...] embodiments of the present invention (90)
In [yet another / an] alternative embodiment (30)	[...] embodiments are merely exemplary of the invention (1)
In alternative embodiments (37)	[...] embodiment of the system of the present invention (1)
Various embodiments (108)	
According to some embodiments (60)	[...] embodiment of the system according to the invention (1)
In the embodiment / embodiments (365)	
Total: 934 occurrences	Total: 749 occurrences

Table 4. Selected Collocations for Embodiment / Embodiments

(2) ***Various embodiments*** *can include a biomass gasifier reactor 200 and receiver configuration that can include various reactor dimensions, shape, and material*. (US Patent 9,150,802)

(3) *In an* ***embodiment of the invention****, the impeller defines upright, angled fins mounted to a base*. (US Patent 8,851,062)

Once again, the USPTO glossary deems another seemingly irrelevant word worthy of definition: *wherein*. The different regulations and rules include this word within the category of clauses whose meaning can limit the property covered by patent claims. Table 5 displays all the limiting clauses considered by MPEP 2111.04 and their corresponding standard interpretation. In our corpus, *wherein* is the most frequent limiting clause. As we can deduce from its standard meaning, this frequency is the result of its lack of patent-subject specificity.

Limiting Clauses	**Standard Meaning in Patents**
Adapted to (149)	Limiting when used in machine claims
Adapted for (22)	Limiting when used in machine claims
Whereby (57)	Not limiting when in *method* claims
Wherein (1,916)	Limiting when giving meaning and purpose to manipulative steps

Table 5. Limiting Clauses (in claim body)

Examples (4) and (5) illustrate how two of these limiting clauses are used.

(4) *An appliance for use in an environment comprising: an appliance housing; an interface* ***adapted to*** *receive power information; a plurality of sensors for sensing environmental conditions; a plurality of controls for controlling operations of the appliance; …* (US Patent 9,010,133)

(5) *The appliance of claim 1* wherein *the appliance includes a fan and the control values include a fan time delay*. (US Patent 9,010,133)

Most Frequent Colligations of *wherein*	
The [constituent of the invention] of claim + [cardinal number], *wherein*	1,095
[constituent of the invention] + according to claim + [cardinal number], *wherein*	263
Wherein + [the/a/said/Ø] + [nominal group] + [further] + comprises	230
Wherein the + [ordinal number] + [nominal group]	155
Wherein the at least + [one/some] + [nominal group]	41
Wherein at least + [a / one / some / cardinal number] + [of the] + [nominal partitive]	38
Wherein the plurality of + [plural nominal group] + [comprise / include]	34
Wherein each of said plurality of [plural nominal group] + [comprises /is configure.]	17
Wherein each of the plurality of + [plural nominal group] + [comprises / provides]	13

Table 6. *Wherein* Colligations

The most frequent colligation forms part of a dependent claim. *Wherein* is explicitly defined in the US patent to introduce limitations to the patent protection, but in 1MEPAC its four most frequent colligations can be used to identify dependent claims [see example (6)]:

(6) *The article of claim 1, **wherein** the container satisfies at least one of the following expedients: (i) exhibits a water vapor transmission rate (WVTR) of less than about 0.3 grams per 100 square inches per 1 day (g/100 in.sup.2/day), as determined by ASTM 1249-06; ...* (US Patent 8,083,064)

The USPTO glossary also includes a series of expressions which can be used to introduce the scope of a patent claim (MPEP 2111.03). Table 7 presents such expressions and their standard interpretation. Examples (7), (8), and (9) illustrate how three of these expressions are typically used.

Transitional Expression			**Standard Meaning in Patents**
	Absolute freq.	Normalized freq. (/10,000 words)	
Characterized by *Characterised by* *Comprising* *Containing* *Including*	17 8 1,085 561 685	0.15 0.07 10.05 5.2 6.35	Are inclusive or open-ended and do not exclude additional, unmentioned elements or method steps
Consisting essentially of	10	0.09	Limits the scope of a claim to the specified materials or steps and those that do not materially affect the basic and novel characteristics of the invention
Consisting of *Consists of*	3 57	0.02 0.52	Excludes any element, step, or ingredient not specified in the claim.
Having	989	9.15	Interpretable according to context.
Composed of	78	0.72	Interpreted either as *consisting of* or as *consisting essentially of* depending on context.

Table 7. Transitional Phrases (immediately after claim preambles)

(7) [...] *a flow deflector **comprising** an edge proximate to the rotating body, the flow deflector guiding the water flow into the curved vanes for allowing operation in various water flow speeds ranging from 1 mile per hour ("MPH") to over 20 MPH; ...* (US Patent 9512816)

(8) *Device according to claim 6, wherein the containment lung has a first end **having** a wider cross-section and a second end **having** an outlet to act as a pressure multiplier*. (US Patent 8,525,365)

(9) *The system of claim 17, wherein the energy controller is configured to control a ventilation system of a garage **containing** the vehicle*. (US Patent 8,872,379)

1MEPAC, as expected, corroborates the hypothesis that transitional expressions are used to set intellectual property limitations both in independent and dependent claims according to what *MPEP 2111.03 Transitional Phrases* establishes (it does not specify any difference whether the claims is dependent or independent).

5.3. *-able/-ible* adjectives

In his guide for patent drafters, Roberts (2007) recommends describing the moving parts of an invention with deverbal adjectives which contain the suffix *-able/-ible*. As Roberts (2007: 86) writes, "it is important to ensure that the claim recites that they are *movable* rather than *moving* as otherwise the claim will not be infringed by a product when it is switched off". Indeed, if one describes the invention when it is functioning, the implication is that the applicants are patenting an invention only in its operating mode. Consequently, a similar description of another invention, in which the components are not moving, would also be patentable.

A search with the TXM concordancer for instances of adjectives ending with the suffix *-ble* yields the following results: 318 different types and a total of 5,893 occurrences.[22] This means that for every 10,000 words one can potentially find 55 instances of such adjectives or approximately the same amount per patent. Table 8 displays the 20 most frequent types of *-ble* adjectives in the corpus.

A Selection of Adjectives with *-ble* Suffix (318 types, 5,893 tokens in total)			
Adjective		Occurrences	Normalized freq. (/10,000 words)
1.	Renewable	446	4
2.	Suitable	371	3
3.	Available	328	3
4.	Possible	324	3
5.	Capable	267	2
6.	Rechargeable	251	2
7.	Able	215	2
8.	Variable	190	2
9.	Combustible	187	1
10.	Rotatable	161	1
11.	Movable	141	1
12.	Programmable	135	1
13.	Desirable	120	1
14.	Flexible	117	1
15.	Acceptable	95	Less than 1
16.	Recyclable	90	Less than 1
17.	Adjustable	84	Less than 1
18.	Operable	82	Less than 1
19.	Uninterruptable	76	Less than 1
20.	Size-adjustable	32	Less than 1

Table 8. List of *-able/-ible* Adjectives

By looking at these adjectives in context, it becomes clear that each one assigns meanings that go beyond those that Roberts (2007) suggested. Examples (10), (11), (12), and (13) below show contexts in which this type of adjectives simply assigns a useful attribute to certain constituents of the patented invention. Examples (14), (15), and (16) present features that cannot be measured

22 The tagging of patent texts seems to generate a higher rate of tag-assignment error than texts with shorter sentences and paragraphs. Moreover, spelling mistakes and the use of mixed spelling standards may generate overrepresented words (the same word appearing as two or more types) and misassigned tags. Most current taggers have an accuracy of over 97% although this value drops to about 90% when tagging unknown words. For an overview see: https://www.aclweb.org/aclwiki/index.php?title=POS_Tagging_(State_of_the_art) (accessed March 27, 2018). Patents contain many words absent from the manually tagged corpora used as reference for tagging new texts.

against any widely accepted technical standard (Arinas Pellón 2012) and calculatedly avoid revealing details of the invention.[23] Only examples (17), (18), and (19) illustrate the rhetorical use as described by Roberts.

(10) [...] *each made from a* ***renewable*** *material, a recycled material, a regrind material, or a mixture thereof.* (US Patent 8,083,064)

(11) [...] *can include an optional small* ***rechargeable*** *or disposable battery and/or storage capacitor* [...] (US Patent 8,851,062)

(12) [...] *circulating pumps and* ***variable*** *speed drives* [...] (US Patent 9,080,789)

(13) *The hot* ***combustible*** *product gases leaving the multiple hearth furnace (activated carbon production reactor) at the top is sent to a combustor/steam boiler*. (US Patent 9,121,606)

(14) *Thus, a* ***suitable*** *amount of the draw solution should directly flow into the buffer chamber 70*. (US Patent 9,474,998)

(15) [...] *the catalyst may be supported in the reactor or may be found as* ***desirable*** *components in the mineral limestone feed*. (US Patent 9,505,998)

(16) *The present disclosure also describes* ***acceptable*** *and preferred growth conditions for hydrogen production*. (US Patent 9,506,084)

(17) *The bays may simply correspond to separately located groupings of related* ***operable*** *components with one another,* [...] (US Patent 9,276,418)

(18) [...] *comprising: an annular housing* ***rotatable*** *about an axis; a permanent magnet freely* ***movable*** *inside said annular housing* [...] (US Patent 8,829,696)

(19) *The solar cooking apparatus is* ***adjustable*** *and, in some embodiments,* ***portable***. (US Patent 9,377,215)

Most adjectives with a *-ble* ending are used for a referential type of description whose purpose is to transmit knowledge (Bal 2009: 46) to the PHOSITA.[24] A small, but prototypically frequent, number of these adjectives (*acceptable*, *desirable*, *possible*, *preferable*, and *suitable*) provide a description vague enough to either claim unforeseen technical details or conveniently hide useful (although not legally essential) details of the invention (Channell 1994: 173-188). Therefore, the strategy proposed by Roberts can be performed with a group of deverbal adjectives (*variable*, *rotatable*, *movable*, *adjustable*, *length-adjustable*, and *size-adjustable*), which can also serve the purpose of describing features of the invention according to functionality. These deverbal adjectives are mostly derived from verbs indicating manner of motion (Levin 1993: 263), which is consistent with Roberts' advice for describing a potential movement of the invention.

5.4. Locating novelty, non-obviousness and usefulness through technical keywords

As indicated in Section 2, patent applications must be convincing in terms of novelty, non-obviousness and usefulness. This objective has to be reached through argumentative strategies which refer specifically to the technology developed in the patent. Therefore, by observing the contextualization of Group B keywords, we can attempt to locate how this reasoning is prototypically carried out in patents. More specifically, we focus on the section known as

23 One reason for this could be that the inventors themselves ignore these details and by their choice of word ensure the property claim of something which may later be discovered by their competitors.

24 Person having ordinary skill in the art, referring to someone with knowledge in the field of a particular patent or invention.

"Background of the Invention", its purpose being to persuade the awarding body of the novelty, non-obviousness, and utility of an invention by resorting to the following two strategies:

a) Identifying the field and scope of the invention patented.

b) Identifying an unsolved (or as yet unsatisfactorily solved) problem in previous inventions.

To identify the patterns used to argue in favour of the patented invention, a mere observation of the keywords used for the query is not sufficient; rather, substantial co-text is necessary. Therefore, the strategy followed consists of reading 10 random sample paragraphs containing the keywords, checking that such paragraphs are in the section of the patent under investigation, testing their automatic identification, and checking anew that the concordances obtained are in the correct section of the patent.

	Absolute Frequencies	**/10,000 Words**
Verbs	Eliminate (78) Improve (258) Lose (42) Overcome (32) Prevent (242) Represent (178) Suffer (18)	0.72 2.39 0.38 0.29 2.24 1.65 0.16
Nouns	Advantage(s) (199) Challenge(s) (28) Concern(s) (29) Disadvantage(s) (20) Drawback(s) (18) Improvement(s) (32) Issue(s) (42) Problem(s) (215) Solution(s) (757)	1.84 0.25 0.26 0.18 0.16 0.29 0.38 1.99 7.01
Adjectives	Common (120) Critical (31) Difficult (45) Inefficient (23) Numerous (39) Significant (116)	1.11 0.28 0.41 0.21 0.36 1.07
Words introducing technical limitations	Associated with (42) Can not/cannot/can't (68) Do not/don't (109) For instance, (167) However (460) (A) [nominal cluster] (is/are) needed (24) There (may be/exists/is) (also) a (clear/growing) need (in/to/for) (38) Without the (97) While (658)	0.38 0.63 1.01 1.54 4.26 0.22 0.35 0.89 6.1

Table 9. Pointers to Limitations in Prior Art

Table 9 displays the most frequent collocates for Group B keywords. These collocates point to forms of semantic prosody that select negative contexts for prior art deficiencies[25] (and positive contexts to highlight desirable characteristics.[26] Any mention of desirable features is presented as currently insufficient, so as to create room for the patented innovation.

25 See examples (20) to (24).

26 See examples (21), (23), and (24).

The three lists below show some of the low-frequency collocates which accompany the keywords in Group B. They are semantically consistent with the higher frequency collocates shown in Table 9:

- Adjectives: *adverse* (4), *complex* (25), *costly* (18), *daunting* (1), *disadvantageous* (3), *excessive* (16), *irreversible* (1), *optimal* (7), *potential* (21), *problematic* (5), *serious* (2), etc.
- Nouns: *benefit(s)* (12), *degradation* (3), *difficulty* / *difficulties* (12), *shortcomings* (3), *shortfall(s)* (2), etc.
- Verbs: *can cause* (5), *fail* (11), *hinder* (4), *offset* / *off-set* (8), *overwhelm* (1), etc.

As examples (20) to (24) illustrate, two difficulties emerge when using concordancers to query for semantic prosodies and preferences in patents. The first one relates to the diversity of lexical items employed to highlight technical deficiencies in prior art. Statistical tools will not show most of them as they occur only occasionally. Therefore, a manual check of the context becomes unavoidable and, consequently, it is deemed to be to some extent unexhaustive. The second problem is related to the context itself: for example, the combination *associated with* occurs a total of 368 times in our corpus, but only 42 of these occurrences are relevant in the context of limitations in prior art. Moreover, some of the collocates need to be seen in context to be interpreted correctly. The words *reliability*, *quality, reliable, consistent*, and *attractive* in example (21) refer to the missing features that the patented invention should include; however, if analysed in a word list with no context, they would usually have a positive connotation.

(20) *Alternative separation schemes 28 have been proposed for separating* ***carbon dioxide*** *from the hydrocarbon streams* ***to avoid*** *the aforementioned* ***challenges associated with*** *packed beds*. (US Patent 9,453,174)

(21) *There are supply reliability, quality, and cost* ***issues associated with*** *this approach such that a more reliable, higher quality and consistent* $\boldsymbol{CO_2}$ *feedstock source would be attractive*. (US Patent 9,272,912)

(22) *In Pennsylvania, the current municipal landfills are projected to become* ***overwhelmed*** *in the near future as the production of* ***gas*** *from shale formations, and thus the* ***waste*** *streams* ***associated with*** *that production, dramatically* ***increases***. (US Patent 8,807,871)

(23) *The achievements of the past notwithstanding, further enhancements to Rankine cycle* ***waste heat*** *recovery systems and methods* ***are needed***. (US Patent 9,260,982)

(24) ***While LED lighting*** *is becoming an attractive option for certain applications, it is not optimal for many applications. Therefore,* ***there is a need for*** *improved LED lighting systems*. (US Patent 9,392,669)

Essentially, patent drafters portray prior art in the "Background of the Invention" section as faulty, insufficient or problematic by collocating the keywords in Group B within the semantic field of 'problems'. This argumentative structure favours the occurrence of verbs such as *overcome*, *eliminate*, and *prevent*, which generates the need for the solution that the patent then describes. Thus, this pattern does not only corroborate the usefulness of the invention, but also its novelty and non-obviousness. Example (24) illustrates the wording of this prototypical reasoning (usually found at the end of that section): objections to current state-of-the-art introduced by *while* and stressing the pressing necessity for a solution by using *there is a need for*.

6. Conclusions

The rationale behind this exploratory and descriptive study lies in the awareness that a deeper understanding of key vocabulary and phraseologies may yield insights into the specificities of a genre such as the patent, whose importance is continuously growing in present-day society.

Patent law discourages accurate disclosure of inventions and instead suggests limiting the amount of information provided and using vagueness, which may lead to "intentional obscurity" (McJohn 2012: 961). Thus, the analysis of authentic patents shows that linguistic strategies are employed to find the essential balance between providing the information necessary for the purposes of achieving the patent and the need to protect one's invention. For instance, linguistic vagueness can guarantee the fulfilment of formal requirements without disclosing precious data.

The considerations drawn may be of use for teaching purposes, especially at institutions specializing in training patent writers and reviewers, but also to develop reading skills of highly-specialised technical texts, both for native and non-native speakers of English. Material created *ad hoc* for learning purposes is often unable to meet the real needs of students, who may then encounter difficulties when confronted with real-world texts which are often characterized by a high level of complexity. This also happens in the case of patents, whose dual nature (technical and legal) contributes the generating difficulty in their comprehension. Instead, the analysis of authentic patents can allow students to empirically discover specific textual patterns which can be employed to disseminate or hide technical information (according to the specific requirements of the different moves of a patent) or to locate novelty, non-obviousness and usefulness through technical words.

In this regard, the use of corpora for understanding patents from a structural and textual perspective can be of particular importance for educational and training purposes. Students will be able to recognize the standardization process imposed by legal constraints in terms of macrostructure, but also in terms of the lexical choices made. Patents can be used as pointers to the consequences of international environmental initiatives as they use legislation-based technological improvements to justify the usefulness of some inventions. This genre is especially suitable for Law and Engineering students to practice reading comprehension that is embedded into regular professional practices. The analysis of the recommendations included in manuals and guides can thus be combined with the observation of authentic texts. The insertion of the production of patents as a genre within technical writing curricula would be a very demanding task, especially at undergraduate levels, as there is a huge amount of expert knowledge needed (of both laws and professional practices) in order to be able to write each type of patent. Nevertheless, the considerations offered in this study may provide guidelines to aid students' understanding of patents and to identify specific language strategies used in order to claim as much intellectual property as possible without violating somebody else's intellectual property rights.

More specialized courses, for example postgraduate students specializing in intellectual property law and with some professional experience, could also make use of analyses of this type in order to make lexical and semantic choices which are in line with the conventions of the genre, based on the corpus analyzed. Instead, undergraduate students would benefit from learning how to read patents, in that they may observe an illustrative example of how the audience, the purpose and the structure influence each other. This activity may also have applications in teaching writing skills, but not necessarily in writing patents themselves. Reading and understanding patents is the first step in understanding the relationship between the genre and the patent community; therefore, reading comprehension as a skill practiced on English for Occupational Purposes courses should not be underestimated.

A close observation of these documents allows readers to be aware of the complex relationships which exist between standardization and specificity and between explicitness and implicitness in technical texts. Given the stability and the conservatism generally associated with patents, specific compositional practices are expected. At the same time, any category of patents may have specific peculiarities, especially from a lexical and semantic perspective. Thus, investigating keyness, and in particular how keyness operates in specialized discourse, contributes to critical and educational studies. More specifically, it allows us to depict the complex network of semantic patterns which typify environmental patents as socially-situated constructs. Indeed, keywords play an important function in helping us not only in the hermeneutic processing of a text in lexico-semantic terms,

but more broadly they are fundamental for knowledge management and for the understanding of the conceptual structuring of a text.

References

Adolphs, Svenja 2006: *Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies*. London: Routledge.

Arinas Pellón, Ismael 2010: How does a patent move? Genre analysis has something to say about it. In Gotti, Maurizio / Williams, Christopher (eds) *Legal Discourse across Languages and Cultures*. Bern: Peter Lang, 313-334.

Arinas Pellón, Ismael 2012: How Vague can your Patent Be? Vagueness Strategies in US Patents. In *HERMES-Journal of Language and Communication in Busines*s 48, 55-74.

Arinas Pellón, Ismael 2014: Interpersonal Patent Relations: Persuasion Pointers to Novelty, Creativity, and Ownership in US Patents. In Breeze, Ruth / Gotti, Maurizio / Sancho Guinda, Carmen (eds) *Interpersonality in Legal Genres.* Bern: Peter Lang.

Bondi, Marina 2010: Perspectives on Keywords and Keyness. In Bondi, Marina / Scott, Mike (eds) *Keyness in Texts.* Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-18.

Brugnoli, Patrizia 2007: Il Testo Giuridico, Apolide in una Tipologia dei Testi? Alcune Riflessioni. In *Revista de Llengua i Dret* 47, 43-59.

Burge, David A. 1999: *Patent and Trademark: Tactics and Practice*. New York: John Wiley & Sons Inc.

Burk, Dan L. / Reyman, Jessica 2014: *Patents as Genre*: A Prospectus. In *Law & Literature* 26/2, 163-190.

Channell, Joanna 1994: *Vague Language*. Oxford: Oxford University Press.

Cole, Paul 2006: *Fundamentals of Patent Drafting.* London: The Chartered Institute of Patent Attorneys.

Ewing, Thomas et al. 2014: *WIPO Patent Drafting Manual. IP Assets Management Series*. World Intellectual Property Organisation. Available at:

http://www.wipo.int/edocs/pubdocs/en/patents/867/wipo_pub_867.pdf (accessed June 20, 2018).

Goźdź-Roszkowski, Stanislaw 2011: P*atterns of Linguistic Variation in American Legal English. A Corpus-based Study*. Frankfurt am Main: Peter Lang Verlag.

Hunston, Susan / Francis, Gill 2000: Pattern Grammar: *A Corpus-Driven Approach to the Lexical Grammar of English.* Amsterdam / Philadelphia: John Benjamins Publishing Company.

Lawson, Veronica 1997: The Terms and Arts of Patentese: Wolves in Sheep's Clothing. In Wright, Sue Ellen / Budin, Gerhard (eds) *Handbook of Terminology*: Volume 1, 171-183. Amsterdam / Philadelphia: John Benjamin's Publishing Company.

Levin, Beth 1993: English Verb Classes and Alternations. Chicago: Chicago University Press.

McJohn, Stephen M. 2012: Patents: Hiding from History. In *Santa Clara High Technology Law Journal* 24, 961-980.

Myers, Greg 1995: From Discovery to Invention: The Writing and Rewriting of Two Patents. In *Social Studies of Science* 25/1, 57-105.

Partington, Alan 1998: Patterns and Meanings: *Using Corpora for English Language Research and Teaching*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Rayson, Paul / Berridge, Damon / Francis, Brian 2004: Extending the Cochran Rule for the Comparison of Word Frequencies. *JADT 2004: 7es Journées Internationales d'Analyse statistique des Données Textuelles.* http://eprints.lancs.ac.uk/12424/1/rbf04_jadt.pdf (accessed June 10, 2018).

Rayson, Paul / Garside, Roger 2000: Comparing Corpora Using Frequency Profiling. WCC '00 *Proceedings of the Workshop on Comparing Corpora 9.* http://www.comp.lancs.ac.uk./~paul/publications/rg_acl2000.pdf (accessed March 10, 2011).

Roberts, Gwilym 2007: *A Practical Guide to Drafting Patents*. London: Sweet & Maxwell Limited.

Scott, Mike 2008: *WordSmith Tools 5.0* Help, Lexical Analysis Software, Liverpool.

Slusky, Ronald D. 2007: Invention Analysis and Claiming: *A Patent Lawyer's Guide*. Chicago, IL: American Bar Association.

Stubbs, Michael 2002: *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell Publishing.

Stubbs, Michael 2009: The Search for Units of Meaning: Sinclair on Empirical Semantics in *Applied Linguistics* 30/1, 115-137.

United States Patent and Trademark Office 2018: *Manual of Patent Examining Procedure*. Ninth Edition, March 2014, Latest Revision January 2018 [R-08.2017], available at: https://www.uspto.gov/web/offices/pac/mpep/index.htm (accessed July 10, 2018).

Wenger, Etienne / McDermott, Richard / Snyder, William M. 2002: Cultivating Communities of Practice: *A Guide to Managing Knowledge*. Boston, MA: Harvard Business Press.